

# MOD-DISCRETE EXPANSIONS

A. D. BARBOUR, E. KOWALSKI, AND A. NIKEGHBALI

**ABSTRACT.** In this paper, we consider approximating expansions for the distribution of integer valued random variables, in circumstances in which convergence in law cannot be expected. The setting is one in which the simplest approximation to the  $n$ 'th random variable  $X_n$  is by a particular member  $R_n$  of a given family of distributions, whose variance increases with  $n$ . The basic assumption is that the ratio of the characteristic function of  $X_n$  and that of  $R_n$  converges to a limit in a prescribed fashion. Our results cover a number of classical examples in probability theory, combinatorics and number theory.

## 1. INTRODUCTION

In a remarkable paper, Hwang (1999) considered sequences of non-negative integer valued random variables  $X_n$ , whose probability generating functions  $f_{X_n}$  satisfy

$$e^{\lambda_n(1-z)} f_{X_n}(z) \rightarrow g(z),$$

for all  $z \in \mathbb{C}$  with  $|z| \leq \eta > 1$ , where the function  $g$  is analytic, and  $\lim_{n \rightarrow \infty} \lambda_n = \infty$ . Under some extra conditions, he exhibits tight bounds on the accuracy of the approximation of the distribution of  $X_n$  by a Poisson distribution with carefully chosen mean, close to  $\lambda_n$ . Independently, motivated by specific examples arising in Random Matrix Theory and number theory, Jacod, Kowalski and Nikeghbali (2008) explored the properties of a related ratio convergence for real valued random variables, namely when the characteristic functions  $\phi_{X_n}$  satisfy

$$e^{-i\theta\beta_n + \theta^2\gamma_n/2} \phi_{X_n}(\theta) \rightarrow \Phi(\theta),$$

locally uniformly in  $\theta$  (in particular, bounds on the error in the approximation of the distribution  $P_{X_n}$  by the normal distribution  $\mathcal{N}(\beta_n, \gamma_n)$  can be simply deduced). Kowalski and Nikeghbali (2009) went on to explore some consequences (and structural aspects in arithmetic cases) of the corresponding uniform limit

$$(1.1) \quad \exp\{\lambda_n(1 - e^{i\theta})\} \phi_{X_n}(\theta) \rightarrow \psi(\theta), \quad 0 < |\theta| \leq \pi,$$

---

2000 *Mathematics Subject Classification.* 62E17; 60F05, 60C05, 60E10, 11N60.

*Key words and phrases.* mod-Poisson convergence; characteristic function; Poisson-Charlier expansion; Erdős-Kac theorem.

Work of EK supported in part by the National Science Foundation under agreement No. DMS-0635607 during a sabbatical stay at the Institute for Advanced Study.

Work of ADB and AN supported in part by Schweizerischer Nationalfonds Projekte Nr. 20-117625/1 (ADB) and Schweizerischer Nationalfonds Projekte Nr. 200021-119970/1 (AN).

for random variables  $X_n$  (usually integer valued), as in Hwang (1999) with the Poisson characteristic function in the ratio. Note that the conditions on the distributions of the  $X_n$  are now much weaker than those of Hwang (1999): for instance, his conditions require the  $X_n$  to take only non-negative values, and to have exponential tails. On the other hand, the probabilistic results that Hwang derives are much more sophisticated. He establishes bounds on the error in his approximations with respect to a number of different metrics, and shows that they are sharp. For instance, for the Kolmogorov and total variation distances, his bounds are typically of order  $O(\lambda_n^{-1})$ , and he also gives the value of the leading asymptotic term in the error.

In this paper, we work with integer valued random variables, and with characteristic function conditions that sharpen (1.1), with the aim of developing approximations of higher order. Our main result, Proposition 2.1, is very simple and explicit. This enables us to dispense with asymptotic settings, and to prove concrete error bounds. As a direct consequence, we are able to deduce a Poisson–Charlier approximation with error of order  $O(\lambda_n^{-(r+1)/2})$ , for any prescribed  $r$ , assuming that Hwang’s conditions hold. Our Poisson–Charlier expansions are derived under more general conditions, in which the  $X_n$  may have only a few finite moments. These are established in Section 3, and simpler, translated Poisson approximations are considered in Section 4.

Hwang (1999) notes that his methods are also applicable to families of distributions other than the Poisson family, and gives examples using the Bessel family. Our approach allows one to derive expansions based on any discrete family of distributions, as shown in Section 5, provided that their characteristic functions satisfy a simple condition, and this without any extra effort. Indeed, the main problem is to identify the higher order terms in the expansions. These turn out to be simply the higher order differences of the basic distribution, leading, for example, to the Charlier polynomial factors in the Poisson case. We discuss some examples, to sums of independent integer valued random variables, to Hwang’s setting and to the Erdős–Kac theorem, in Section 6.

**Remark.** We recall the motivation behind the terminology (mod-gaussian, mod-poisson, and here mod-discrete): the simplest example leading to limits like (say) (1.1) is when  $X_n = P_n + Y$  where  $P_n$  has Poisson distribution  $\text{Po}(\lambda_n)$  and is independent of  $Y$ , where  $\psi(\theta)$  is the characteristic function of  $Y$ . Thus the sequence converges to  $Y$  “modulo Poisson variables”.

## 2. THE BASIC ESTIMATE

We frame our approximations in terms of three distances between (signed) measures  $\mu$  and  $\nu$  on the integers: the point metric

$$d_{\text{loc}}(\mu, \nu) := \sup_{j \in \mathbb{Z}} |\mu\{j\} - \nu\{j\}|,$$

the Kolmogorov distance

$$d_K(\mu, \nu) := \sup_{j \in \mathbb{Z}} |\mu\{(-\infty, j]\} - \nu\{(-\infty, j]\}|,$$

and the total variation norm

$$\|\mu - \nu\| := \sum_{j \in \mathbb{Z}} |\mu\{j\} - \nu\{j\}| = 2 \sup_{A \subset \mathbb{Z}} |\mu(A) - \nu(A)|.$$

Other metrics could also be treated using our results. Our conditions are expressed in terms of characteristic functions, defined, for a finite signed measure  $\sigma$  on  $\mathbb{Z}$ , by  $\phi_\sigma(\theta) := \sum_{j \in \mathbb{Z}} e^{ij\theta} \sigma\{j\}$ , for  $|\theta| \leq \pi$ . The essence of our argument is the following simple result, linking the closeness of the signed measures to the closeness of their characteristic functions, when these have a common factor involving a ‘large’ parameter  $\rho$ .

**Proposition 2.1.** *Let  $\mu$  and  $\nu$  be finite signed measures on  $\mathbb{Z}$ , with characteristic functions  $\phi_\mu$  and  $\phi_\nu$  respectively. Suppose that  $\phi_\mu = \psi_\mu \chi$  and  $\phi_\nu = \psi_\nu \chi$ , where, for some  $\gamma_1, \gamma_2, \rho, t > 0$ ,*

$$|\psi_\mu(\theta) - \psi_\nu(\theta)| \leq \gamma_1 |\theta|^t \quad \text{and} \quad |\chi(\theta)| \leq \gamma_2 e^{-\rho\theta^2} \quad \text{for all } |\theta| \leq \pi.$$

*Then, writing  $\gamma = \gamma_1 \gamma_2$ , there are explicit constants  $\alpha_{1t}$  and  $\alpha_{2t}$  such that*

1.  $\sup_{j \in \mathbb{Z}} |\mu\{j\} - \nu\{j\}| \leq \alpha_{1t} \gamma (\rho \vee 1)^{-(t+1)/2};$
2.  $\sup_{a \leq b \in \mathbb{Z}} |\mu\{[a, b]\} - \nu\{[a, b]\}| \leq \alpha_{2t} \gamma (\rho \vee 1)^{-t/2}.$

*Proof.* For any  $j \in \mathbb{Z}$ , the Fourier inversion formula gives

$$(2.1) \quad \mu\{j\} - \nu\{j\} = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-ij\theta} (\psi_\mu(\theta) - \psi_\nu(\theta)) \chi(\theta) d\theta,$$

from which our assumptions imply directly that

$$|\mu\{j\} - \nu\{j\}| \leq \frac{1}{2\pi} \int_{-\pi}^{\pi} \gamma |\theta|^t \exp\{-\rho\theta^2\} d\theta.$$

For  $\rho \leq 1$ , we thus have

$$|\mu\{j\} - \nu\{j\}| \leq \frac{1}{2\pi} \int_{-\pi}^{\pi} \gamma |\theta|^t d\theta \leq \frac{\pi^t \gamma}{t+1} =: \beta_{1t} \gamma.$$

For  $\rho \geq 1$ , it is immediate that

$$|\mu\{j\} - \nu\{j\}| \leq \frac{\gamma}{2\pi} \left( \frac{1}{\sqrt{2\rho}} \right)^{t+1} \int_{-\infty}^{\infty} |y|^t e^{-y^2/2} dy \leq \beta'_{1t} \gamma \rho^{-(t+1)/2},$$

with  $\beta'_{1t} := 2^{-(t+1)/2} m_t / \sqrt{2\pi}$ ; here,  $m_t$  denotes the  $t$ -th absolute moment of the standard normal distribution. Setting  $\alpha_{1t} := \max\{\beta_{1t}, \beta'_{1t}\}$ , this proves part 1. The second part is similar, adding (2.1) over  $a \leq j \leq b$ , and estimating

$$\frac{|e^{-ia\theta} - e^{-i(b+1)\theta}|}{|1 - e^{-i\theta}|} \leq \frac{\pi}{|\theta|}, \quad |\theta| \leq \pi.$$

This gives part 2, with

$$\alpha_{2t} := \max\{2^{-t/2} m_{t-1} \sqrt{\pi/2}, \pi^t/t\}.$$

In particular, the second part bounds the distance between the two measures in the Kolmogorov distance. We shall principally be concerned with taking  $\mu$  to be the distribution of a random variable  $X$ ; we allow  $\nu$  to be a signed measure largely for reasons of technical convenience.

For some applications, a slight weakening of the conditions in Proposition 2.1 is useful. The following result is proved in exactly the same way as before.

**Proposition 2.2.** *Let  $\mu$  and  $\nu$  be finite signed measures on  $\mathbb{Z}$ , with characteristic functions  $\phi_\mu$  and  $\phi_\nu$  respectively. Suppose that  $\phi_\mu = \psi_\mu \chi$  and  $\phi_\nu = \psi_\nu \chi$ , where, for some  $\theta_0, \gamma, \varepsilon, \eta, \rho' > 0$  and for positive pairs  $\gamma_m, t_m$ ,  $1 \leq m \leq M$ , we have*

$$|\psi_\mu(\theta) - \psi_\nu(\theta)| \leq \sum_{m=1}^M \gamma_m |\theta|^{t_m} + \varepsilon \quad \text{and} \quad |\chi(\theta)| \leq \gamma e^{-\rho\theta^2}, \quad 0 \leq |\theta| \leq \theta_0;$$

$$|\phi_\mu(\theta) - \phi_\nu(\theta)| \leq \eta, \quad \theta_0 < |\theta| \leq \pi.$$

Then, with notation as for Proposition 2.1, we have

$$1. \quad \sup_{j \in \mathbb{Z}} |\mu\{j\} - \nu\{j\}| \leq \sum_{m=1}^M \gamma_m \gamma \alpha_{1t_m} (\rho \vee 1)^{-(t_m+1)/2} + \tilde{\alpha}_1 \gamma \varepsilon + \tilde{\alpha}_2 \eta;$$

$$2. \quad \sup_{a_0 \leq a \leq b \leq b_0} |\mu\{[a, b]\} - \nu\{[a, b]\}|$$

$$\leq \sum_{m=1}^M \gamma_m \gamma \alpha_{2t_m} (\rho \vee 1)^{-t_m/2} + (b_0 - a_0 + 1)(\tilde{\alpha}_1 \gamma \varepsilon + \tilde{\alpha}_2 \eta),$$

where

$$\tilde{\alpha}_1 := \left( \frac{\theta_0}{\pi} \wedge \frac{1}{2\sqrt{\pi\rho}} \right); \quad \tilde{\alpha}_2 := \left( 1 - \frac{\theta_0}{\pi} \right).$$

The presence of the factor  $(b_0 - a_0 + 1)$  in the second bound means that a direct bound on the Kolmogorov distance between the signed measures  $\mu$  and  $\nu$  is not immediately visible. The following corollary is however easily deduced.

**Corollary 2.3.** *Under the conditions of Proposition 2.2,*

$$d_K(\mu, \nu) \leq \inf_{a \leq b} (\varepsilon_{ab}^{(K)} + |\mu|\{(-\infty, a) \cup (b, \infty)\} + |\nu|\{(-\infty, a) \cup (b, \infty)\});$$

$$\|\mu - \nu\| \leq \inf_{a \leq b} (\varepsilon_{ab}^{(1)} + |\mu|\{(-\infty, a) \cup (b, \infty)\} + |\nu|\{(-\infty, a) \cup (b, \infty)\}),$$

where

$$\varepsilon_{ab}^{(K)} := \sum_{m=1}^M \gamma_m \gamma \alpha_{2t_m} (\rho \vee 1)^{-t_m/2} + (b - a + 1)(\tilde{\alpha}_1 \gamma \varepsilon + \tilde{\alpha}_2 \eta);$$

$$\varepsilon_{ab}^{(1)} := (b - a + 1) \left\{ \sum_{m=1}^M \gamma_m \gamma \alpha_{1t_m} (\rho \vee 1)^{-(t_m+1)/2} + (\tilde{\alpha}_1 \gamma \varepsilon + \tilde{\alpha}_2 \eta) \right\}.$$

If also  $\mu$  is a probability measure, then

$$d_K(\mu, \nu) \leq \inf_{a \leq b} (1 - \nu\{[a, b]\} + 2\varepsilon_{ab}^{(K)} + |\nu|\{(-\infty, a) \cup (b, \infty)\}).$$

*Proof.* The inequality for the total variation norm is immediate. For the Kolmogorov distance, by considering the possible positions of  $x$  in relation to  $a < b$ , we have

$$\begin{aligned} & |\mu\{(-\infty, x]\} - \nu\{(-\infty, x]\}| \\ & \leq \sup_{y < a} |\mu\{(-\infty, y]\} - \nu\{(-\infty, y]\}| + \sup_{a \leq y \leq b} |\mu\{[a, y]\} - \nu\{[a, y]\}| \\ & \quad + \sup_{y > b} |\mu\{(b, y]\} - \nu\{(b, y]\}| \\ & \leq |\mu\{(-\infty, a) \cup (b, \infty)\}| + |\nu\{(-\infty, a) \cup (b, \infty)\}| + \varepsilon_{ab}^{(K)}. \end{aligned}$$

If  $\mu$  is a probability measure, we have

$$|\mu\{(-\infty, a) \cup (b, \infty)\}| = 1 - \mu\{[a, b]\} \leq 1 - \nu\{[a, b]\} + \varepsilon_{ab}^{(K)}. \quad \square$$

Under slightly stronger conditions than those of Proposition 2.1, a much neater total variation bound can be deduced.

**Proposition 2.4.** *Let  $\mu$  and  $\nu$  be finite signed measures on  $\mathbb{Z}$ , with characteristic functions  $\phi_\mu = \psi_\mu \chi$  and  $\phi_\nu = \psi_\nu \chi$  respectively, where  $\chi(\theta) := \gamma_2 e^{-u(\theta)}$  for some  $\gamma_2 > 0$ , and  $u(0) = 0$ . Suppose now that  $u$  and the difference  $d_{\mu\nu} := \psi_\mu - \psi_\nu$  are both twice differentiable, that  $u'(0) = d'_{\mu\nu}(0) = 0$  and that, for some  $\gamma_1, \gamma_2, \gamma_3 > 0$ ,  $\rho \geq 1$  and  $t \geq 2$ ,*

$$|d''_{\mu\nu}(\theta)| \leq \gamma_1 |\theta|^{t-2}, \quad |u''(\theta)| \leq \gamma_3 \rho \quad \text{and} \quad u(\theta) \geq \rho \theta^2 \quad \text{for all } |\theta| \leq \pi.$$

*Then, writing  $\gamma = \gamma_1 \gamma_2$ , there is a constant  $\alpha_3 := \alpha_3(t, \gamma_3)$  such that*

$$\|\mu - \nu\| \leq \alpha_3 \gamma (\rho \vee 1)^{-t/2}.$$

*Proof.* First, the assumptions on  $d_{\mu\nu}$  and  $u$  give

$$(2.2) \quad \begin{aligned} |d'_{\mu\nu}(\theta)| &\leq \frac{\gamma_1}{t-1} |\theta|^{t-1}; \quad |d_{\mu\nu}(\theta)| \leq \frac{\gamma_1}{t(t-1)} |\theta|^t; \\ |u'(\theta)| &\leq \gamma_3 \rho |\theta|. \end{aligned}$$

In particular, for  $|j| \leq \lceil \sqrt{\rho} \rceil$ , we can apply part 1 of Proposition 2.1, which gives

$$(2.3) \quad |\mu\{j\} - \nu\{j\}| \leq \frac{\alpha_1 t \gamma}{t(t-1)} (\rho \vee 1)^{-(t+1)/2}.$$

For the remaining  $j$ , integrating the Fourier inversion formula (2.1) twice by parts gives

$$(2.4) \quad \begin{aligned} \mu\{j\} - \nu\{j\} = & -\frac{1}{2\pi j^2} \int_{-\pi}^{\pi} e^{-ij\theta} \left( d''_{\mu\nu}(\theta) - 2d'_{\mu\nu}(\theta)u'(\theta) + \right. \\ & \left. d_{\mu\nu}(\theta)\{(u'(\theta))^2 - u''(\theta)\} \right) \chi(\theta) d\theta. \end{aligned}$$

Substituting the bounds from (2.2) into (2.4) gives

$$\begin{aligned} & |\mu\{j\} - \nu\{j\}| \\ & \leq \frac{1}{2\pi j^2} \int_{-\pi}^{\pi} \gamma \left\{ |\theta|^{t-2} + \frac{2\gamma_3\rho}{t-1} |\theta|^t + \frac{\gamma_3\rho}{t(t-1)} |\theta|^t (1 + \gamma_3\rho\theta^2) \right\} e^{-\rho\theta^2} d\theta \\ & \leq \frac{\rho}{j^2} \gamma \beta_3(t, \gamma_3) \rho^{-(t+1)/2}, \end{aligned}$$

after some calculation, where, with  $m_t$  as in Proposition 2.1,

$$\beta_3(t, \gamma_3) := \frac{m_{t-2}}{4t 2^{t/2} \sqrt{\pi}} \{4t + (2t+1)\gamma_3 + (t+1)\gamma_3^2\}.$$

Hence

$$\sum_{|j| > \lceil \sqrt{\rho} \rceil} |\mu\{j\} - \nu\{j\}| \leq 2\gamma \beta_3(t, \gamma_3) \rho^{-t/2},$$

and the proposition follows directly, with  $\alpha_3(t, \gamma_3) := 2\beta_3(t, \gamma_3) + \frac{5\alpha_1 t}{t(t-1)}$ .  $\square$

### 3. POISSON-CHARLIER EXPANSIONS

Suppose first that  $X$  is an integer valued random variable having characteristic function  $\phi_X$  of the form  $\phi_X(\theta) = \psi(\theta)p_\lambda(\theta)$ ,  $|\theta| \leq \pi$ , where  $p_\lambda$  denotes the characteristic function of the Poisson distribution  $\text{Po}(\lambda)$  with mean  $\lambda$ . Underlying our considerations is an unspecified asymptotic setting in which  $\lambda$  is large and  $\psi$  is thought of as (almost) fixed, but we do not need to make direct use of this. We now assume in addition that, for some  $r \in \mathbb{N}_0$  and for some  $K_{r\delta} > 0$ ,  $0 < \delta \leq 1$ ,

$$(3.1) \quad |\psi(\theta) - \psi_r(\theta)| \leq K_{r\delta} |\theta|^{r+\delta}, \quad |\theta| \leq \pi,$$

where

$$(3.2) \quad \psi_r(\theta) := \sum_{l=0}^r a_l (i\theta)^l$$

is a polynomial of degree  $r$  with real coefficients  $a_l$ , thus implying that  $a_0 = 1$ . If  $\psi$  is itself the characteristic function of a probability measure, this assumption roughly corresponds to assuming that the measure has (at least)  $r$  finite moments. Alternatively, we could assume that

$$(3.3) \quad |\psi(\theta) - \tilde{\psi}_r(\theta)| \leq K_{r\delta} |\theta|^{r+\delta}, \quad |\theta| \leq \pi,$$

where

$$(3.4) \quad \tilde{\psi}_r(\theta) := \sum_{l=0}^r \tilde{a}_l (e^{i\theta} - 1)^l,$$

again with real coefficients  $\tilde{a}_l$  and  $\tilde{a}_0 = 1$ . If  $r = 0$ , and thus  $\psi_0(\theta) = 1$  for all  $\theta$ , we could now immediately use (3.1) in conjunction with Proposition 2.1 to approximate the distribution of  $X$  by the Poisson distribution  $\text{Po}(\lambda)$ , with an error in Kolmogorov distance of order  $\lambda^{-\delta/2}$ ; note that

$$(3.5) \quad |p_\lambda(\theta)| = \exp\{-\lambda(1 - \cos \theta)\} \leq e^{-\rho\theta^2}, \quad |\theta| \leq \pi,$$

with  $\rho := 2\pi^{-2}\lambda$ .

We now want to go further, and use (3.1) with higher values of  $r$  to justify more sophisticated approximations with a higher order of accuracy. In order to do so, we need to find ‘nice’ signed measures  $\nu_r$ , whose characteristic functions are at least as close to  $\psi_r(\theta)p_\lambda(\theta)$  and  $\tilde{\psi}_r(\theta)p_\lambda(\theta)$  as  $\psi(\theta)p_\lambda(\theta)$  is. Now  $\tilde{\psi}_r(\theta)p_\lambda(\theta)$  is itself the characteristic function of a signed measure, which we can then take as our choice of  $\nu_r$ . To identify  $\nu_r$ , observe that, if  $\phi_\mu$  is the characteristic function of a signed measure  $\mu$ , then  $(e^{i\theta} - 1)^l \phi_\mu(\theta)$  is the characteristic function of the  $l$ ’th difference  $\Delta^l \mu$  of  $\mu$ :

$$\Delta^l \mu\{j\} := \sum_{k=0}^l \binom{l}{k} (-1)^k \mu\{j - l + k\}.$$

For  $\mu$  the Poisson distribution, this yields the Poisson–Charlier signed measures:

$$(3.6) \quad \tilde{\psi}_r(\theta)p_\lambda(\theta) = \sum_{l=0}^r \tilde{a}_l (e^{i\theta} - 1)^l p_\lambda(\theta)$$

is the characteristic function of the signed measure  $\nu = \nu_r(\lambda; \tilde{a}_1, \dots, \tilde{a}_r)$  on  $\mathbb{N}_0$  defined by

$$(3.7) \quad \nu\{j\} := \text{Po}(\lambda)\{j\} \left\{ 1 + \sum_{l=1}^r (-1)^l \tilde{a}_l C_l(j; \lambda) \right\},$$

where

$$(3.8) \quad C_l(j; \lambda) := \sum_{k=0}^l (-1)^k \binom{l}{k} \binom{j}{k} k! \lambda^{-k}$$

denotes the  $l$ -th Charlier polynomial (Chihara 1978, (1.9), p. 171).

Note that, if  $\binom{j}{k}$  is replaced by  $j^k/k!$  in (3.8), one obtains the binomial expansion of  $(1 - j/\lambda)^l$ . As this suggests, the values of  $C_l(j; \lambda)$  are in fact small for  $j$  near  $\lambda$  if  $\lambda$  is large:

$$(3.9) \quad |C_l(j; \lambda)| \leq 2^{l-1} \{|1 - j/\lambda|^l + (l/\sqrt{\lambda})^l\}$$

(Barbour & Čekanavičius 2002, Lemma 6.1). (3.9) thus implies that the  $l$ -th term in the sum in (3.7) has total variation norm at most  $|\tilde{a}_l| c_l \lambda^{-l/2}$ , for a universal constant  $c_l$ . It also implies that, in any interval of the form  $|j - \lambda| \leq c\sqrt{\lambda}$ , which is where the probability mass of  $\text{Po}(\lambda)$  is mostly to be found, the correction to the Poisson measure  $\text{Po}(\lambda)$  is of uniform relative order  $O(\lambda^{-l/2})$ . Indeed, the Chernoff inequalities for  $Z \sim \text{Po}(\lambda)$  can be expressed in the form

$$(3.10) \quad \max\{\mathbb{P}[Z > \lambda(1 + \delta)], \mathbb{P}[Z < \lambda(1 - \delta)]\} \leq \exp\{-\lambda\delta^2/2(1 + \delta/3)\},$$

for  $0 < \delta \leq 1$  (Chung & Lu 2006, Theorem 3.2). Since also, from (3.8),

$$|C(j; \lambda)| \leq (1 + j/\lambda)^l \leq 2^l \quad \text{if } 0 \leq j \leq \lambda,$$

and since

$$\binom{j}{k} k! \lambda^{-k} \frac{e^{-\lambda} \lambda^j}{j!} = \frac{e^{-\lambda} \lambda^{j-k}}{(j-k)!} \leq \frac{e^{-\lambda} \lambda^{j-l}}{(j-l)!}$$

if  $0 \leq k \leq l$  and  $j \geq l + \lambda$ , it follows that, for any  $l \geq 0$ , we have

$$\sum_{j=0}^m |C_l(j; \lambda)| \text{Po}(\lambda)\{j\} \leq 2^l \mathbb{P}[Z \leq m] \leq 2^l \exp\{-(\lambda - m)^2/3\lambda\}$$

for  $m \leq \lambda$ , and

$$\sum_{j \geq m} |C_l(j; \lambda)| \text{Po}(\lambda)\{j\} \leq 2^l \mathbb{P}[Z \geq m - l] \leq 2^l \exp\{-(m - l - \lambda)^2/3\lambda\},$$

for  $\lambda + l \leq m \leq 2\lambda + l$ .

Writing  $|\nu|$  to denote the absolute measure associated with  $\nu$ , it thus follows that

$$(3.11) \quad \begin{aligned} |\nu|\{[0, m]\} &\leq \bar{A}_r e^{-(\lambda - m)^2/3\lambda}, & 0 \leq m \leq \lambda; \\ |\nu|\{[m, \infty)\} &\leq \bar{A}_r e^{-(m - r - \lambda)^2/3\lambda}, & \lambda + r \leq m \leq 2\lambda, \end{aligned}$$

where  $\bar{A}_r := 1 + \sum_{l=1}^r 2^l |\tilde{a}_l|$ , demonstrating concentration of measure for  $\nu$  on a scale of  $\sqrt{\lambda}$  around  $\lambda$ . Moreover, it can be deduced from (3.9) that there exists a positive constant  $d = d(\tilde{a}_1, \dots, \tilde{a}_r)$  such that  $\nu\{j\} \geq 0$  for  $|j - \lambda| \leq d\lambda$ , and it follows from (3.11) that  $|\nu|\{j: |j - \lambda| > d\lambda\} = O(e^{-\alpha\lambda})$  for some  $\alpha > 0$ . Since also  $\nu\{\mathbb{N}_0\} = 1$ , it thus follows that, even if  $\nu$  is formally a signed measure, it differs from a probability only on a set of measure exponentially small with  $\lambda$ .

Thus, if (3.3) holds, it follows that  $X$  has characteristic function  $\psi(\theta)p_\lambda(\theta)$  and  $\nu := \nu_r(\lambda; \tilde{a}_1, \dots, \tilde{a}_r)$  has characteristic function  $\phi_\nu = \tilde{\psi}_r(\theta)p_\lambda(\theta)$ , and that the conditions of Proposition 2.1 are satisfied with  $\mu = P_X$ ,  $t = r + \delta$ ,  $\kappa = K_{r\delta}$  and  $\rho = 2\pi^{-2}\lambda$ , this last from (3.5). If, instead, we are given the inequality (3.1), we can write  $e^{i\theta} - 1 = i\theta \sum_{s \geq 0} (i\theta)^s / (s + 1)!$ , and equate the coefficients of  $(i\theta)^j$  in (3.2) with those for  $1 \leq j \leq r$  in (3.6), giving  $\tilde{a}_1, \dots, \tilde{a}_r$  implicitly in terms of  $a_1, \dots, a_r$ :

$$(3.12) \quad a_j = \sum_{l=1}^j \tilde{a}_l \sum_{(s_1, \dots, s_l) \in S_{j-l}} \prod_{t=1}^l \frac{1}{(s_t + 1)!},$$

where  $S_m := \{(s_1, \dots, s_l): \sum_{t=1}^l s_t = m\}$ . With this choice of  $\tilde{a}_1, \dots, \tilde{a}_r$ , it follows that  $\nu = \nu_r(\lambda; \tilde{a}_1, \dots, \tilde{a}_r)$  has characteristic function  $\phi_\nu$  satisfying

$$(3.13) \quad |\psi_r(\theta) - \phi_\nu(\theta)| \leq \Gamma_r |\theta|^{r+1}, \quad |\theta| \leq \pi,$$

for  $\Gamma_r := \Gamma_r(a_1, \dots, a_r)$ . Hence, in this case, we obtain

$$(3.14) \quad |\psi(\theta) - \phi_\nu(\theta)| \leq (K_{r\delta} + G_{r\delta}) |\theta|^{r+\delta}, \quad |\theta| \leq \pi,$$

with  $G_{r\delta} := \Gamma_r \pi^{1-\delta}$ , and the conditions of Proposition 2.1 are satisfied with  $\mu = P_X$ ,  $t = r + \delta$ ,  $\gamma = K_{r\delta} + G_{r\delta}$  and  $\rho = 2\pi^{-2}\lambda$ . Thus, if either (3.1) or (3.3) is satisfied, a signed measure from the family  $\nu_r(\lambda; b_1, \dots, b_r)$  can be found, which approximates the probability measure  $P_X$  in the sense implied by Proposition 2.1. These measures are themselves rather explicit perturbations of the Poisson distribution  $\text{Po}(\lambda)$ .

We summarize these considerations in the following theorem, which is deduced directly from Proposition 2.1. Note that we shall later be primarily concerned with applications in which  $K_{r\delta}$  and  $G_{r\delta}$  are not small, and in



which therefore  $\lambda$  must be big, if our bounds are to be useful. However, for the sake of completeness, we phrase our bounds in a form which also allows for accuracy of approximation if  $K_{r\delta} + G_{r\delta}$  is small.

**Theorem 3.1.** *Let  $X$  be a random variable on  $\mathbb{Z}$  with distribution  $P_X$ . Suppose that its characteristic function  $\phi_X$  is of the form  $\psi p_\lambda$ , where  $p_\lambda(\theta)$  is the characteristic function of the Poisson distribution  $\text{Po}(\lambda)$  with mean  $\lambda$ . Suppose also that (3.1) is satisfied, for some  $r \in \mathbb{N}_0$  and  $\delta \geq 0$ . Let  $\nu_r := \nu_r = \nu_r(\lambda; \tilde{a}_1, \dots, \tilde{a}_r)$  be as in (3.7), with  $\tilde{a}_1, \dots, \tilde{a}_r$  given implicitly by (3.12). Then, writing  $t = r + \delta$ , we have*

$$\begin{aligned} 1. \quad d_{\text{loc}}(P_X, \nu_r) &:= \sup_{j \in \mathbb{Z}} |P_X\{j\} - \nu_r\{j\}| \\ &\leq \alpha'_{1t}(K_{r\delta} + G_{r\delta})(\lambda \vee 1)^{-(t+1)/2}; \\ 2. \quad d_K(P_X, \nu_r) &:= \sup_{l \in \mathbb{Z}} |P_X\{(-\infty, l]\} - \nu_r\{[0, l]\}| \\ &\leq \alpha'_{2t}(K_{r\delta} + G_{r\delta})(\lambda \vee 1)^{-t/2}, \end{aligned}$$

with

$$\begin{aligned} \alpha'_{1t} &:= \alpha_{1t}(\pi^2/2)^{(t+1)/2}, \quad \alpha'_{2t} = \alpha_{2t}(\pi^2/2)^{t/2}, \\ G_{r\delta} &:= \Gamma(a_1, \dots, a_r)\pi^{1-\delta}. \end{aligned}$$

If (3.1) is replaced by (3.3), the corresponding bounds hold with  $G_{r\delta} = 0$ .

Theorem 3.1 enables one to deduce simple bounds for other measures of the distance between  $P_X$  and  $\nu$ . For instance, for the total variation norm, with judicious choice of  $m_1$  and  $m_2$ , we can use part 1 to bound

$$(3.15) \quad \sum_{j=m_1+1}^{m_2-1} |P_X\{j\} - \nu\{j\}| \leq (m_2 - m_1 - 1) \sup_{j \in \mathbb{N}_0} |P_X\{j\} - \nu\{j\}|,$$

and then (3.11) and part 2 to take care of the remaining tail probabilities:

$$\begin{aligned} (3.16) \quad \sum_{j \leq m_1} |P_X\{j\} - \nu\{j\}| &\leq P_X\{(-\infty, m_1]\} + |\nu\{[0, m_1]\}| \\ &\leq \sup_{l \in \mathbb{N}_0} |P_X\{(-\infty, l]\} - \nu\{[0, l]\}| + 2|\nu\{[0, m_1]\}|, \end{aligned}$$

and

$$\begin{aligned} (3.17) \quad \sum_{j \geq m_2} |P_X\{j\} - \nu\{j\}| &\leq P_X\{[m_2, \infty)\} + |\nu\{[m_2, \infty)\}| \\ &\leq \sup_{l \in \mathbb{N}_0} |P_X\{(-\infty, l]\} - \nu\{[0, l]\}| + 2|\nu\{[m_2, \infty)\}|. \end{aligned}$$

This gives the following theorem.

**Theorem 3.2.** *Suppose that the conditions of Theorem 3.1 are satisfied, with (3.14) holding. If  $K_{r\delta} + G_{r\delta} < 1$ , there is a constant  $\alpha_{4t}$  such that*

$$(3.18) \quad \|P_X - \nu\| \leq \alpha_{4t}(K_{r\delta} + G_{r\delta})(\lambda \vee 1)^{-t/2} \max\{1, \sqrt{|\log(K_{r\delta} + G_{r\delta})|}, \sqrt{\log(\lambda + 1)}\};$$

if  $K_{r\delta} + G_{r\delta} \geq 1$  and  $\lambda^{(r+1)/2} \geq K_{r\delta} + G_{r\delta}$ , then there is a constant  $\alpha_{5t}$  such that

$$(3.19) \quad \|P_X - \nu\| \leq \alpha_{5t}(K_{r\delta} + G_{r\delta})\lambda^{-t/2} \max\{1, \sqrt{\log(\lambda + 1)}\}.$$

*Proof.* For  $K_{r\delta} + G_{r\delta} < 1$  and  $\lambda \geq 1$ , we use both parts of (3.11), with

$$m_1 := \lfloor \lambda - c_{r\lambda} \sqrt{\lambda \log(\lambda + 1)} \rfloor \quad \text{and} \quad m_2 := \lceil \lambda + r + c_{r\lambda} \sqrt{\lambda \log(\lambda + 1)} \rceil,$$

where  $\lfloor x \rfloor \leq x \leq \lceil x \rceil$  denote the integers closest to  $x$ , obtaining

$$|\nu|(\{[0, m_1] \cup [m_2, \infty)\}) \leq 2B_r(\lambda + 1)^{-c_{r\lambda}^2/3} \leq 2B_r(K_{r\delta} + G_{r\delta})(\lambda + 1)^{-(r+1)/2},$$

if  $c_{r\lambda}^2 := 3(r+1)/2 + |\log(K_{r\delta} + G_{r\delta})|/\log(\lambda + 1)$ . Hence, from (3.15)–(3.17), it follows that

$$\begin{aligned} \|P_X - \nu\| &\leq \{2c_{r\lambda} \sqrt{\lambda \log(\lambda + 1)} + r + 2\} \alpha'_{1t}(K_{r\delta} + G_{r\delta}) \lambda^{-(t+1)/2} \\ &\quad + 2\alpha'_{2t}(K_{r\delta} + G_{r\delta}) \lambda^{-t/2} + 4B_r(K_{r\delta} + G_{r\delta}) \lambda^{-(r+1)/2}, \end{aligned}$$

so that

$$\begin{aligned} \|P_X - \nu\| &\leq \beta_{3t}(K_{r\delta} + G_{r\delta}) \lambda^{-t/2} \\ &\quad \max\{1, \sqrt{\log(1/(K_{r\delta} + G_{r\delta}))}, \sqrt{\log(\lambda + 1)}\}, \end{aligned}$$

with  $\beta_{3t} := \alpha'_{1t}\{\sqrt{6(r+1)} + r + 4\} + 2\alpha'_{2t} + 4B_r$ .

For  $K_{r\delta} + G_{r\delta} < 1$  and  $\lambda < 1$ , we take  $m_2 := \lceil \lambda + r + \sqrt{3|\log(K_{r\delta} + G_{r\delta})|} \rceil$  in (3.11), giving

$$|\nu|(\{[m_2, \infty)\}) \leq B_r(K_{r\delta} + G_{r\delta}),$$

and

$$\begin{aligned} \|P_X - \nu\| &\leq (r + 2 + \sqrt{3|\log(K_{r\delta} + G_{r\delta})|}) \alpha'_{1t}(K_{r\delta} + G_{r\delta}) \\ &\quad + 2\alpha'_{2t}(K_{r\delta} + G_{r\delta}) + 2B_r(K_{r\delta} + G_{r\delta}), \end{aligned}$$

so that

$$\|P_X - \nu\| \leq \beta'_{3t}(K_{r\delta} + G_{r\delta}) \max\{1, \sqrt{|\log(K_{r\delta} + G_{r\delta})|}, \sqrt{\log(\lambda + 1)}\},$$

with  $\beta'_{3t} := \alpha'_{1t}\{r + 4\} + \alpha_{2t} + 2B_r$ . Then (3.18) follows, with  $\alpha_{4t} := \max\{\beta_{3t}, \beta'_{3t}\}$ .

For  $\lambda^{t/2} \geq K_{r\delta} + G_{r\delta} \geq 1$ , we take  $m_1 := \lfloor \lambda - c_r \sqrt{\lambda \log(\lambda + 1)} \rfloor$  and  $m_2 := \lceil \lambda + r + c_r \sqrt{\lambda \log(\lambda + 1)} \rceil$ , with  $c_r := \sqrt{3t/2}$ , giving

$$|\nu|(\{[0, m_1] \cup [m_2, \infty)\}) \leq 2B_r(\lambda + 1)^{-t/2}$$

Using (3.15) and (3.17), it follows that

$$\begin{aligned} \|P_X - \nu\| &\leq \{2c_r \sqrt{\lambda \log(\lambda + 1)} + r + 2\} \alpha'_{1t}(K_{r\delta} + G_{r\delta}) \lambda^{-(t+1)/2} \\ &\quad + 2\alpha'_{2t}(K_{r\delta} + G_{r\delta}) \lambda^{-t/2} + 4B_r \lambda^{-t/2} \\ &\leq \alpha_{5t}(K_{r\delta} + G_{r\delta}) \lambda^{-t/2} \max\{1, \sqrt{\log(\lambda + 1)}\}, \end{aligned}$$

with  $\alpha_{5t} = \alpha'_{1t}(\sqrt{6(r+1)} + r + 2) + 2\alpha'_{2t} + 4B_r$ .

Note that if  $K_{r\delta} + G_{r\delta} \geq 1$  and  $\lambda^{t/2} < K_{r\delta} + G_{r\delta}$ , one cannot hope to get a useful bound from Theorem 3.1. If  $\lambda \geq 1$ , the error bound for the individual probabilities is then at least of size  $\lambda^{-1/2}$ , which is the same size as many of the probabilities themselves. If  $\lambda < 1$ , the bound on the individual

probabilities is of size comparable to 1.  $\square$

**Remark.** Under the extra conditions that  $\psi$  is twice differentiable and that either (3.1) or (3.3) holds with  $r \geq 2$ , Proposition 2.4 shows that the factor  $\sqrt{\log(\lambda + 1)}$  can in fact be dispensed with. Note that, to satisfy the conditions of the proposition, it is necessary to take  $\chi(\theta) := \exp\{e^{i\theta} - 1 - i\theta\}$ , to get  $u'(0) = 0$ .

Sometimes it is convenient, for simplicity, to use parameters in the expansions that are not those emerging naturally from the proofs. The following theorem shows that such alterations can easily be allowed for.

**Theorem 3.3.** *Suppose that*

$$\phi_\mu := p_\lambda A; \quad \phi_{\nu^{(1)}} := p_\lambda A'; \quad \phi_{\nu^{(2)}} := p_{\lambda'} A,$$

with  $A(\theta) := 1 + \sum_{l=1}^r a_l \theta^l$ ,  $A'(\theta) := 1 + \sum_{l=1}^r a'_l \theta^l$  and with  $\lambda > \lambda'$ . Then, with  $\rho := 2\pi^{-2}\lambda$ ,  $\rho' := 2\pi^{-2}\lambda'$  and  $a_0 := 1$ ,

$$\begin{aligned} d_{\text{loc}}(\mu, \nu^{(1)}) &\leq \sum_{l=1}^r \alpha_{1l} |a_l - a'_l| (\rho \vee 1)^{-(l+1)/2}; \\ d_K(\mu, \nu^{(1)}) &\leq \sum_{l=1}^r \alpha_{2l} |a_l - a'_l| (\rho \vee 1)^{-l/2}; \\ d_{\text{loc}}(\mu, \nu^{(2)}) &\leq (\lambda - \lambda') \sum_{l=1}^{r+1} \alpha_{1l} |a_{l-1}| (\rho' \vee 1)^{-(l+1)/2}; \\ d_K(\mu, \nu^{(2)}) &\leq (\lambda - \lambda') \sum_{l=1}^{r+1} \alpha_{2l} |a_{l-1}| (\rho' \vee 1)^{-l/2}. \end{aligned}$$

*Proof.* For the comparison between  $\mu$  and  $\nu^{(1)}$ , we have

$$|A(\theta) - A'(\theta)| \leq \sum_{l=1}^r |a_l - a'_l| |\theta|^l, \quad 0 < |\theta| \leq \pi,$$

and Proposition 2.2 completes the proof. For that between  $\mu$  and  $\nu^{(2)}$ , note that  $p_\lambda = p_{\lambda-\lambda'} p_{\lambda'}$ , and that, for  $\lambda > \lambda'$  and  $0 < |\theta| \leq \pi$ ,

$$|p_{\lambda-\lambda'}(\theta) - 1| |A(\theta)| \leq (\lambda - \lambda') |\theta| \left\{ 1 + \sum_{l=1}^r |a_l| |\theta|^l \right\},$$

from which and Proposition 2.2 the remaining results follow.  $\square$

#### 4. POISSON APPROXIMATION

The measures  $\nu_r$  considered above are very explicit. Nevertheless, it is even neater to have approximation in terms of a Poisson distribution, where possible. Clearly, if (3.1) holds for any  $r = r_0$ ,  $\delta = \delta_0$ , then it holds with  $r = 0$  and  $\psi_0(\theta) = 1$  for all  $\theta$ , with the exponent  $r + \delta$  replaced by  $\delta_0$  if  $r_0 = 0$  and by 1 if  $r_0 \geq 1$ , with  $K_0$  depending on  $K_{r_0}$  and on  $a_1, \dots, a_{r_0}$ . Theorem 3.1

then gives approximation by  $\text{Po}(\lambda)$  with accuracy in Kolmogorov distance of order  $O(\lambda^{-t_0/2})$ , for  $t_0 = \min\{1, r_0 + \delta_0\}$ .

However, if  $r_0 \geq 1$ , one can also write

$$\psi(\theta)p_\lambda(\theta) = \hat{\psi}(\theta)p_{\lambda'}(\theta),$$

for any  $\lambda' > 0$ , where

$$\hat{\psi}(\theta) := \psi(\theta) \exp\{(\lambda - \lambda')(e^{i\theta} - 1)\}.$$

Taking  $\lambda' - \lambda = a_1$  now gives a bound

$$|\hat{\psi}(\theta) - 1| \leq K'_1 |\theta|^{t_1},$$

of the form (3.1), with  $t_1 = \min\{r_0 + \delta_0, 2\}$ . Hence, Theorem 3.1 implies the following approximation.

**Corollary 4.1.** *If  $X$  has characteristic function  $\phi_X(\theta) = \psi(\theta)p_\lambda(\theta)$  such that (3.1) is satisfied with  $r \geq 1$ , then we have*

1.  $d_{\text{loc}}(P_X, \text{Po}(\lambda')) \leq \alpha_{1t} \gamma(\rho' \vee 1)^{-(t+1)/2};$
2.  $d_K(P_X, \text{Po}(\lambda')) \leq \alpha_{2t} \gamma(\rho' \vee 1)^{-t/2},$

where  $\lambda' = \lambda + a_1$ ,  $t := \min\{2, r + \delta\}$  and  $\rho' = 2\pi^{-2}\lambda'$ .

The parameter  $\lambda'$  is chosen to make the Poisson mean  $\lambda'$  equal to the mean  $\lambda + a_1$  of  $X$ . This choice of the Poisson parameter improves the rate, in the asymptotic sense that, if  $a_1, \dots, a_r$  and  $K_{r\delta}$  remain bounded but  $\lambda \rightarrow \infty$ , and if  $r + \delta \geq 2$ , then the approximation error for Kolmogorov distance is of order  $O(\lambda^{-1})$ , as opposed to the rate of order  $O(\lambda^{-1/2})$  in general obtained when approximating by  $\text{Po}(\lambda)$ .

Analogously, fitting the second moment as well (if it is finite) can lead to further improvement. If Poisson approximation is still the aim, the easiest way to proceed is to consider translating the random variable  $X$ , and approximating  $X - m$  by a Poisson instead: now one would wish to fix

$$\lambda' = \text{Var } X = \mathbb{E}X - m.$$

This works well if  $\langle \text{Var } X - \mathbb{E}X \rangle = 0$ , where  $\langle x \rangle$  denotes the fractional part of  $x$ , but fails otherwise, since  $X - m$  only remains integer valued if  $m$  is itself an integer. For general  $X$ , we therefore use an average of two adjacent Poisson probabilities to approximate  $P_X\{j\}$ . The details are as follows.

Suppose that (3.1) is satisfied with  $r \geq 2$ :  $\phi_X(\theta) = \psi(\theta)p_\lambda(\theta)$  with

$$|\psi(\theta) - \psi_r(\theta)| \leq K_{r\delta} |\theta|^{r+\delta},$$

where  $\psi_r(\theta) = \sum_{j=0}^r a_j (i\theta)^j$ . For  $m \in \mathbb{Z}$  and  $0 \leq p \leq 1$ , define the probability measure  $Q_{\lambda'mp}$  by

$$(4.1) \quad Q_{\lambda'mp}\{j\} := p \text{Po}(\lambda')\{j - m - 1\} + (1 - p) \text{Po}(\lambda')\{j - m\},$$

having characteristic function  $q_{\lambda'mp}$  given by

$$(4.2) \quad q_{\lambda'mp}(\theta) := e^{im\theta} (1 + p(e^{i\theta} - 1)) p_{\lambda'}(\theta).$$

For  $p = 0$ ,  $Q$  has the distribution of  $Z' + m$ , where  $Z' \sim \text{Po}(\lambda')$ ; for  $p = 1$ ,  $Q$  has the distribution of  $Z' + m + 1$ ; for  $0 < p < 1$ ,  $Q$  is a mixture of these two distributions. Thus the family of distributions  $Q_{\lambda'mp}$  can be interpreted

as a natural generalization of the usual translated Poisson family, in which the translation is not restricted to the integers, but may take any real value. Then we can equivalently write  $\phi_X(\theta) = \bar{\psi}(\theta)q_{\lambda' mp}(\theta)$ , with

$$(4.3) \quad \bar{\psi}(\theta) := \psi(\theta) \exp\{(\lambda - \lambda')(e^{i\theta} - 1) - im\theta\} \{1 + p(e^{i\theta} - 1)\}^{-1}.$$

In the expansion of  $\bar{\psi}(\theta)$ , the coefficients of  $\theta$  and  $\theta^2$  are equal to zero if

$$\mathbb{E}X = \lambda + a_1 = m + \lambda' + p \quad \text{and} \quad \text{Var } X = \lambda + (2a_2 - a_1^2) = \lambda' + p(1 - p).$$

Then  $m \in \mathbb{Z}$ ,  $0 \leq p < 1$  and  $\lambda'$  satisfy these two equations if

$$(4.4) \quad \begin{aligned} m &:= \lfloor a_1 - (2a_2 - a_1^2) \rfloor; \quad p^2 := \langle a_1 - (2a_2 - a_1^2) \rangle; \\ \lambda' &:= \lambda + (2a_2 - a_1^2) - p(1 - p), \end{aligned}$$

and it then follows that

$$(4.5) \quad |\bar{\psi}(\theta) - 1| \leq \gamma |\theta|^t,$$

for suitable choice of  $\gamma$  depending on  $a_1 \dots a_r$  and  $K_{r\delta}$ , with  $t = \min\{3, r + \delta\}$ . Since also, from (4.2) and (3.5),

$$|q_{\lambda' mp}(\theta)| \leq |p_{\lambda'}(\theta)| \leq e^{-\rho' \theta^2},$$

with  $\rho' = 2\pi^{-2}\lambda'$ , the conditions of Proposition 2.1 are satisfied with  $\chi = q_{\lambda' mp}$ , yielding the following corollary.

**Corollary 4.2.** *If  $X$  has characteristic function  $\phi_X(\theta) = \psi(\theta)p_\lambda(\theta)$  such that (3.1) is satisfied with  $r \geq 2$ , then, for  $\lambda'$ ,  $m$  and  $p$  defined as in (4.4) and for  $t := \min\{3, r + \delta\}$ , we have translated Poisson approximation of the form*

1.  $d_{\text{loc}}(P_X, Q_{\lambda' mp}) \leq \alpha_{1t} \gamma (\rho' \vee 1)^{-(t+1)/2};$
2.  $d_K(P_X, Q_{\lambda' mp}) \leq \alpha_{2t} \gamma (\rho' \vee 1)^{-t/2},$

where  $\rho' = 2\pi^{-2}\lambda'$  and  $\gamma$  is as in (4.5). If (3.1) is replaced by (3.3), then one takes  $a_1 := \tilde{a}_1$  and  $a_2 := \tilde{a}_2 + \frac{1}{2}\tilde{a}_1$  in (4.4), to determine  $\lambda'$ ,  $m$  and  $p$ .

In particular, if  $a_1, \dots, a_r$  and  $K_{r\delta}$  remain bounded but  $\lambda \rightarrow \infty$ , and if  $r + \delta \geq 3$ , then  $t = 3$  and the order of approximation in Kolmogorov distance is of order  $O(\lambda^{-3/2})$ .

## 5. MORE GENERAL EXPANSIONS

We now consider cases in which the role of the Poisson family  $\text{Po}(\lambda)$  is replaced by that of another family of probability distributions  $R_\lambda$ ,  $\lambda \geq 1$ , on the integers. We shall assume that, for  $Z_\lambda \sim R_\lambda$ ,  $\mu(\lambda) := \mathbb{E}Z_\lambda$  and  $\sigma^2(\lambda) := \text{Var } Z_\lambda$  exist, and are both continuous functions of  $\lambda$ , with  $\sigma^2(\lambda)$  increasing to infinity with  $\lambda$ . Suppose also that there exist  $c > 0$  and  $h(\lambda)$  such that

$$(5.1) \quad \inf_{\lambda \geq 1} \inf_{0 < |\theta| \leq \pi} \frac{1}{\theta^2 h(\lambda)} \{-\log |r_\lambda(\theta)|\} \geq c,$$

where  $r_\lambda$  is the characteristic function of  $R_\lambda$ . Clearly, if (5.1) is satisfied, one could take

$$h(\lambda) := \inf_{0 < |\theta| \leq \pi} \frac{1}{\theta^2} \{-\log |r_\lambda(\theta)|\}$$

and  $c = 1$ , or else maybe  $h(\lambda) := \sigma^2(\lambda)$  with  $c$  to be determined, but it may also be more convenient to choose some other, simpler form. Then, much as in Section 3, we can consider approximating the distribution of a random variable  $X$  with characteristic function  $\phi_X := \psi(\theta)r_\lambda(\theta)$  by that of a signed measure  $\nu_r = \nu_r(R_\lambda; \tilde{a}_1, \dots, \tilde{a}_r)$  with characteristic function

$$\phi_{\nu_r}(\theta) := \tilde{\psi}_r(\theta)r_\lambda(\theta) := \sum_{l=0}^r \tilde{a}_l(e^{i\theta} - 1)^l r_\lambda(\theta),$$

(as usual,  $\tilde{a}_0 = 1$ ). As in the Poisson case,  $\nu_r$  is just the linear combination  $\sum_{l=0}^r \tilde{a}_l D^l R_\lambda$  of the differences  $D^l R_\lambda$  of the probability measure  $R_\lambda$ . Approximation of the characteristic functions could be expressed either as

$$(5.2) \quad |\psi(\theta) - \tilde{\psi}_r(\theta)| \leq K_{r\delta}|\theta|^{r+\delta}, \quad |\theta| \leq \pi,$$

for real coefficients  $\tilde{a}_l$  and for  $r \in \mathbb{N}_0$ ,  $0 < \delta \leq 1$ , or as

$$(5.3) \quad |\psi(\theta) - \psi_r(\theta)| \leq K_{r\delta}|\theta|^{r+\delta}, \quad |\theta| \leq \pi,$$

where  $\psi_r(\theta)$  is as in (3.2), in which case the corresponding coefficients  $\tilde{a}_l$  can be deduced from (3.12). These considerations lead to the following theorem, following directly from Proposition 2.1.

**Theorem 5.1.** *Let  $X$  be a random variable on  $\mathbb{Z}$  with distribution  $P_X$ . Suppose that its characteristic function  $\phi_X$  is of the form  $\psi R_\lambda$ , where  $R_\lambda$  is as above. Suppose also that (5.2) is satisfied, for some  $r \in \mathbb{N}_0$  and  $\delta \geq 0$ . Then, writing  $t = r + \delta$ , we have*

$$\begin{aligned} 1. \quad d_{\text{loc}}(P_X, \nu_r) &\leq \alpha_{1t} K_{r\delta} (\rho \vee 1)^{-(t+1)/2}; \\ 2. \quad d_K(P_X, \nu_r) &\leq \alpha_{2t} K_{r\delta} (\rho \vee 1)^{-t/2}, \end{aligned}$$

with  $\rho := c h(\lambda)$ ,  $\alpha_{1t}$  and  $\alpha_{2t}$  as in Proposition 2.1, and

$$\nu_r = \nu_r(R_\lambda; \tilde{a}_1, \dots, \tilde{a}_r).$$

If (5.2) is replaced by (5.3), the corresponding bounds hold with  $K_{r\delta}$  replaced by  $K_{r\delta} + G_{r\delta}$ , with  $G_{r\delta} := \Gamma_r \pi^{1-\delta}$  and  $\Gamma_r$  as in (3.13).

As in Section 4, one may prefer to approximate with a suitably translated member of the family  $\{R_\lambda, \lambda \geq 1\}$ , rather than with a signed measure  $\nu_r$ . The corresponding family of distributions  $Q_{mp}(R_\lambda)$ , for  $m \in \mathbb{Z}$  and  $0 \leq p \leq 1$ , is given by

$$(5.4) \quad Q_{mp}(R_\lambda)\{j\} := p R_\lambda\{j - m - 1\} + (1 - p) R_\lambda\{j - m\},$$

having characteristic function  $q_{mp}^{(R_\lambda)}$  given by

$$(5.5) \quad q_{mp}^{(R_\lambda)}(\theta) := e^{im\theta} (1 + p(e^{i\theta} - 1)) r_\lambda(\theta).$$

Once again, the trick is to find  $\lambda'$ ,  $m$  and  $p$  so that the mean and variance of  $X$  and of the distribution  $Q_{mp}(R_\lambda)$  are matched.

If (5.3) is satisfied with  $r \geq 2$ , matching mean and variance implies that we need

$$\begin{aligned} \mathbb{E}X &= \mu(\lambda) + a_1 = m + \mu(\lambda') + p; \\ (5.6) \quad \text{Var } X &= \sigma^2(\lambda) + (2a_2 - a_1^2) = \sigma^2(\lambda') + p(1 - p), \end{aligned}$$

where the coefficients  $a_1$  and  $a_2$  are as in (3.2). These equations have a solution, as long as  $\text{Var } X \geq \sigma^2(1) + 1/4$ , obtained as follows. For  $0 \leq p \leq 1$ , let  $\lambda(p)$  be defined to be the solution of the equation  $\sigma^2(\lambda(p)) = \text{Var } X - p(1-p)$ , noting that  $\lambda(0) = \lambda(1)$ . Choose  $m^* := \lfloor \mathbb{E}X - \mu(\lambda(0)) \rfloor$ . Then the continuous function

$$f(p) := \mathbb{E}X - \mu(\lambda(p)) - m^* - p$$

satisfies  $f(0) \geq 0 > f(1)$ , so that there exists a  $p^*$  such that  $f(p^*) = 0$ . Then the choice  $\lambda' = \lambda(0)$ ,  $m^*$  and  $p^*$  satisfies (5.6), as desired.

**Corollary 5.2.** *If  $X$  has characteristic function  $\phi_X(\theta) = \psi(\theta)R_\lambda(\theta)$  and if (5.3) is satisfied with  $r \geq 2$ , then, for  $\lambda'$ ,  $m$  and  $p$  solving (5.6) and for  $t := \min\{3, r + \delta\}$ , we have translated  $R_\lambda$ -approximation of the form*

$$\begin{aligned} 1. \quad d_{\text{loc}}(P_X, Q_{mp}) &\leq \alpha_{1t} \gamma(\rho' \vee 1)^{-(t+1)/2}; \\ 2. \quad d_K(P_X, Q_{mp}) &\leq \alpha_{2t} \gamma(\rho' \vee 1)^{-t/2}, \end{aligned}$$

where  $\rho' = ch(\lambda')$  and for suitable choice of  $\gamma$ .

The most natural application of the above theorem is to mod-compound Poisson approximation. For  $\lambda > 0$  and for  $\mu$  a probability distribution on  $\mathbb{Z}$ , let  $\text{CP}(\lambda, \mu)$  denote the distribution of the sum  $Y := \sum_{j \in \mathbb{Z} \setminus \{0\}} j Z_j$ , where  $Z_j$ ,  $j \neq 0$ , are independent, and  $Z_j \sim \text{Po}(\lambda \mu_j)$ . Then, if  $\mu_1 > 0$ , the characteristic function of  $Y$  is of the form  $R_\lambda := \zeta_\lambda p_{\lambda_1}$ , where  $\zeta_\lambda$  is the characteristic function of  $\sum_{j \in \mathbb{Z} \setminus \{0,1\}} j Z_j$  and  $\lambda_1 = \lambda \mu_1$ . Thus, for the purposes of applying Theorem 5.1 and Corollary 5.2,  $\rho$  can be taken to be  $2\pi^{-2}\lambda_1$ . Corollary 5.2, for instance, then gives conditions under which translated compound Poisson distribution can be achieved, with approximation at rate  $O(\lambda^{-3/2})$ .

These considerations apply as long as  $\mu_1 > 0$ , and could also be invoked if  $\mu_{-1} > 0$ . If  $\mu_1 = \mu_{-1} = 0$ , there is then no factor of the form  $p_\lambda$  to guarantee that, for some  $\rho > 0$ , the characteristic function  $\phi_Y$  of  $Y$  (corresponding to the characteristic function  $\chi$  of Proposition 2.1) satisfies  $|\phi_Y(\theta)| \leq \exp\{-\rho\theta^2\}$  for all  $|\theta| \leq \pi$ . Some additional aperiodicity condition needs to be satisfied, if the family  $\{\text{CP}(\lambda, \mu), \lambda \geq 1\}$  is to satisfy (5.1). Indeed, if  $Y = 2Z$  where  $Z \sim \text{Po}(\lambda)$ , and if  $W \sim \text{Be}(1/2)$  is independent of  $Y$ , it is not true that the distribution of  $Y + W$  is close to that of  $Y$  in total variation, even though  $|\phi_{Y+W}(\theta) - \phi_Y(\theta)| \leq K_0|\theta|$ .

## 6. APPLICATIONS

**6.1. A single convolution.** The most obvious application of the above results arises when  $\phi_X = \psi p_\lambda$  and  $\psi$  is itself the characteristic function of a probability distribution on the integers. In this case,  $X$  is the sum of two independent random variables, one of them with the  $\text{Po}(\lambda)$  distribution, and the situation is probabilistically very simple. For example, we could take  $\psi$  to be the characteristic function of a random variable  $Y_s$  with

$$\mathbb{P}[Y_s = j] = s! s \sum_{j \geq 1} \frac{1}{j(j+1) \dots (j+s)}$$

for some integer  $s \geq 1$ . Calculation shows that  $Y_s$  has characteristic function

$$\psi(\theta) = 1 + s \sum_{l=1}^{s-1} \frac{(1 - e^{-i\theta})^l}{s-l} - s(1 - e^{-i\theta})^s \log(1 - e^{i\theta}),$$

and that (3.3) holds with  $r = s - 1$  and any  $\delta < 1$  if

$$\tilde{a}_0 = 1; \quad \tilde{a}_j = \sum_{l=1}^j (-1)^{j-l} \frac{s}{s-l} \binom{j-1}{l-1}, \quad 1 \leq j \leq s-1.$$

Hence, if  $X = Z + Y_s$ , where  $Z \sim \text{Po}(\lambda)$  is independent of  $Y_s$ , then the theorems in Sections 3 and 4 can be applied, provided that  $s$  is large enough; in particular, a translated Poisson approximation can be applied with accuracy of order  $O(\lambda^{-3/2+\varepsilon})$  for any  $\varepsilon > 0$  if  $s = 3$  (in which case  $X$  has finite second moment), and of order  $O(\lambda^{-3/2})$  if  $s \geq 4$ . Similar considerations apply to the approximation of  $X = Z - Y_s$ .

**6.2. Sums of independent random variables.** Let  $X_1, \dots, X_n$  be independent integer valued random variables, and let  $S_n$  denote their sum. In contexts in which a central limit approximation to the distribution of  $S_n$  would be appropriate, the classical Edgeworth expansion (see, e.g., Petrov 1975, Chapter 5) is unwieldy, because  $S_n$  is confined to the integers. As an alternative, Barbour and Čekanavičius (2002) give a Poisson–Charlier expansion, for  $S_n$  ‘centered’ so that its mean and variance are almost equal, with an error bound expressed in the total variation norm. Here, we show that such an expansion can be justified by the techniques of this paper.

Assume that each of the  $X_j$  has finite  $(r+1+\delta)$ ’th moment, with  $r \geq 1$ , and define

$$(6.1) \quad A^{(r)}(w) := 1 + \sum_{l \geq 2} \tilde{a}_l^{(r)} w^l = \exp \left\{ \sum_{l=2}^{r+1} \frac{\kappa_l w^l}{l!} \right\},$$

where  $\kappa_l := \kappa_l(S_n)$  and  $\kappa_l(X)$  denotes the  $l$ ’th factorial cumulant of the random variable  $X$ . Then the approximation that we establish is to the Poisson–Charlier signed measure  $\nu_r$  with

$$(6.2) \quad \nu_r\{j\} := \text{Po}(\lambda)\{j\} \left\{ 1 + \sum_{l=2}^{L_r} (-1)^l \tilde{a}_l^{(r)} C_l(j; \lambda) \right\},$$

where  $L_r := \max\{1, 3(r-1)\}$ , and where  $\lambda := \mathbb{E}S_n$ ;  $\nu_r$  has characteristic function

$$(6.3) \quad \phi_{\nu_r} := p_\lambda(\theta) \tilde{A}^{(r)}(\theta),$$

where

$$(6.4) \quad \tilde{A}^{(r)}(\theta) := 1 + \sum_{l=2}^{L_r} \tilde{a}_l^{(r)} (e^{i\theta} - 1)^l.$$

We need two further quantities involving the  $X_j$ :

$$(6.5) \quad K^{(n)} := \left| \sum_{j=1}^n \kappa_2(X_j) \right|$$



and

$$(6.6) \quad p_j := 1 - d_{TV}(\mathcal{L}(X_j), \mathcal{L}(X_{j+1})).$$

**Theorem 6.1.** *Suppose that there are constants  $K_l$ ,  $1 \leq l \leq r+1$ , such that, for each  $j$ ,*

$$|\kappa_l(X_j)| \leq K_l, \quad 2 \leq l \leq r+1; \quad \mathbb{E}|X_j|^{r+1+\delta} \leq K_1^{r+1+\delta}.$$

*Suppose also that  $p_j \geq p_0 > 0$  for all  $j$ , and that  $\lambda \geq n\lambda_0$ . Then*

$$d_K(\mathcal{L}(S_n), \nu_r) \leq G(K_1, \dots, K_{r+1}, K^{(n)}, p_0^{-1}, \lambda_0^{-1}) n^{-(r-1+\delta)/2},$$

*for a function  $G$  that is bounded on compact sets.*

**Remark.** For asymptotics in  $n$ , with triangular arrays of variables, the error is of order  $O(n^{-(r-1+\delta)/2})$  when  $\lambda_0$  and  $p_0$  are bounded away from zero, and  $K_1, \dots, K_{r+1}$  and  $K^{(n)}$  remain bounded. The requirements on  $\lambda_0$  and  $p_0$  can often be achieved by grouping the random variables appropriately, though attention then has to be paid to the consequent changes in the  $K_l$ . The final condition can always be satisfied with  $K^{(n)} \leq 1$ , by replacing the  $X_j$  by translates, where necessary. For more discussion, we refer to Barbour and Čekanavičius (2002). The above conditions are designed to cover sums of independent random variables, each of which has non-trivial variance, has uniformly bounded  $(r+1+\delta)$ 'th moment, and whose distribution overlaps with its unit translate.

*Proof.* We check the conditions of Proposition 2.2. First, in view of (6.6), we can write

$$\mathbb{E}(e^{i\theta X_j}) = \frac{1}{2}p_j(e^{i\theta} + 1)\phi_{1j}(\theta) + (1-p_j)\phi_{2j}(\theta),$$

where both  $\phi_{1j}$  and  $\phi_{2j}$  are characteristic functions. Hence we have

$$|\mathbb{E}(e^{i\theta X_j})| \leq 1 - p_j + p_j \cos(\theta/2) \leq 1 - p_j \theta^2/4\pi, \quad 0 \leq |\theta| \leq \pi.$$

Hence  $\phi_\mu(\theta) := \mathbb{E}(e^{i\theta S_n})$  satisfies

$$(6.7) \quad |\phi_\mu(\theta)| \leq \exp\{-np_0\theta^2/4\pi\}, \quad 0 \leq |\theta| \leq \pi.$$

On the other hand, from the additivity of the factorial cumulants, we have

$$|\kappa_l(S_n)| \leq nK_l, \quad 3 \leq l \leq r+1,$$

with  $|\kappa_2(S_n)| \leq K^{(n)}$  from (6.5). From (6.1), we thus deduce the bound  $|\tilde{a}_l^{(r)}| \leq c_l n^{\lfloor l/3 \rfloor}$ , for  $c_l = c_l(K^{(n)}, K_3, \dots, K_{r+1})$ ,  $l \geq 1$ . Hence

$$(6.8) \quad |\phi_{\nu_r}(\theta)| \leq \exp\{-2n\lambda_0\theta^2/\pi^2\} c' n^{\lfloor L_r/3 \rfloor} \leq \exp\{-n\lambda_0\theta^2/\pi^2\} c'',$$

for  $c'' = c''(K^{(n)}, K_3, \dots, K_{r+1})$ , and we can take  $\eta := Ce^{-n\rho'\theta_0^2}$  in Proposition 2.2, for

$$\rho' = \min\{\lambda_0/\pi^2, p_0/4\pi\}$$

and a suitable  $C = C(K^{(n)}, K_3, \dots, K_{r+1})$ . The choice of  $\theta_0$  we postpone for now.

For  $|\theta| \leq \theta_0$ , we take  $\chi(\theta) := p_\lambda(\theta)$ , and check the approximation of

$$\phi_\mu(\theta) \exp\{-\lambda(e^{i\theta} - 1)\} = \mathbb{E}\{(1+w)^{S_n}\} e^{-w\mathbb{E}S_n}$$

by  $\tilde{A}^{(r)}(\theta)$  as a polynomial in  $w := e^{i\theta} - 1$ . We begin with the inequality

$$\begin{aligned} \left| (1+w)^s - \sum_{l=0}^{r+1} \frac{w^l}{l!} s_{(l)} \right| &\leq \frac{|s_{(r+2)}|}{(r+2)!} |w|^{r+2} \wedge 2 \frac{|s_{(r+1)}|}{(r+1)!} |w|^{r+1} \\ &\leq \frac{|s_{(r+1)}|}{(r+2)!} |w|^{r+1+\delta} \{|s| + r + 1\}^\delta \{2(r+2)\}^{1-\delta}, \end{aligned}$$

derived using Taylor's expansion, true for any  $s \in \mathbb{Z}$  and  $0 < \delta \leq 1$ , where  $s_{(l)} := s(s-1)\dots(s-l+1)$ . Hence, for each  $j$ , we have

$$(6.9) \quad \left| \mathbb{E} \{ (1+w)^{X_j} \} - \sum_{l=0}^{r+1} \frac{\mathbb{E} \{ (X_j)_{(l)} \}}{l!} w^l \right| \leq c_{r,\delta} |\theta|^{r+1+\delta} (K_1 + K_1^{r+1+\delta}),$$

for a universal constant  $c_{r,\delta}$ . Then, writing

$$Q_{r+1}^{(s)}(w; X) := \exp \left\{ \sum_{l=s}^{r+1} \kappa_l(X) w^l / l! \right\},$$

and using the differentiation formula in Petrov (1975, p. 170), we have

$$\begin{aligned} \left| Q_{r+1}^{(1)}(w; X_j) - \sum_{l=0}^{r+1} w^l \mathbb{E} \left( \frac{X_j}{l} \right) \right| &\leq \frac{|\theta|^{r+2}}{(r+2)!} \sup_{|\theta'| \leq \theta_0} \left| \frac{d^{r+2}}{dz^{r+2}} Q_{r+1}^{(1)}(z; X_j) \right|_{z=e^{i\theta'}-1} \\ (6.10) \quad &\leq |\theta|^{r+2} c(K_1, \dots, K_{r+1}), \end{aligned}$$

for a suitable function  $c$  and for all  $|\theta| \leq \pi$ . Combining these estimates, we deduce that, for  $w = e^{i\theta} - 1$  and for all  $|\theta| \leq \pi$ ,

$$(6.11) \quad \left| \mathbb{E} \{ (1+w)^{X_j} \} e^{-\mathbb{E} X_j w} - Q_{r+1}^{(2)}(w; X_j) \right| \leq k_1 |\theta|^{r+1+\delta},$$

where  $k_1 = k_1(K_1, \dots, K_{r+1})$ .

Now a standard inequality shows that, for  $u_j := \prod_{l=1}^j x_l \prod_{l=j+1}^n y_l$ , for complex  $x_l, y_l$  with  $y_l \neq 0$  and  $|x_l/y_l - 1| \leq \varepsilon_l$ , then

$$(6.12) \quad |u_n - u_0| \leq |u_0| \left\{ \prod_{s=1}^{n-1} (1 + \varepsilon_s) \right\} \sum_{l=1}^n \varepsilon_l.$$

Taking  $x_j := \mathbb{E} \{ (1+w)^{X_j} \} e^{-\mathbb{E} X_j w}$  and  $y_j := Q_{r+1}^{(2)}(w; X_j)$ , (6.11) shows that we can take  $\varepsilon_l := \varepsilon := k_1 |\theta|^{r+1+\delta} e^M$  for each  $l$ , with

$$M := \exp \left\{ \sum_{l=2}^{r+1} K_l / l! \right\},$$

provided that  $|\theta| \leq \theta_0 \leq 1$ . Choosing  $\theta_0 := n^{-1/3}$  then ensures that  $(1 + \varepsilon)^n$  is suitably bounded, and (6.12) yields

$$(6.13) \quad \left| \mathbb{E} \{ (1+w)^{S_n} \} e^{-w \mathbb{E} S_n} - Q_{r+1}^{(2)}(w; S_n) \right| \leq k_2 n |\theta|^{r+1+\delta},$$

for  $k_2 = k_2(K^{(n)}, K_1, \dots, K_{r+1})$ , since

$$|Q_{r+1}^{(2)}(w; S_n)| \leq \exp \{ |\kappa_2(S_n)| \theta_0^2 / 2 \} \exp \left\{ \sum_{l=3}^{r+1} n K_l |\theta_0|^l / l! \right\}$$

is bounded for  $\theta_0 = n^{-1/3}$ , in view of (6.5).

The remaining step is to note that, for  $w = e^{i\theta} - 1$ ,

$$(6.14) \quad \left| Q_{r+1}^{(2)}(w; S_n) - \tilde{A}^{(r)}(\theta) \right| \leq \frac{|\theta|^{L_r+1}}{(L_r+1)!} \sup_{|\theta'| \leq \theta_0} \left| \frac{d^{L_r+1}}{dz^{L_r+1}} Q_{r+1}^{(2)}(z; S_n) \right|_{z=e^{i\theta'}-1},$$

where the right hand side is at most  $k_3 n^{r-1} |\theta|^{L_r+1} (1 + n|\theta|^2)$  in  $|\theta| \leq n^{-1/3}$ , with  $k_3 = k_3(K^{(n)}, K_1, \dots, K_{r+1})$ . Here, we use the facts that  $|\kappa_2(S_n)|$  is bounded by  $K^{(n)}$ , and that each  $\kappa_l(S_n)$  for  $l \geq 3$ , for which we have only the weak bound  $nK_l$ , occurs associated with the power  $w^l$  in the exponent of  $Q_{r+1}^{(2)}(w; S_n)$ . Combining this with (6.13), we have established that for  $|\theta| \leq n^{-1/3}$ , we have

$$(6.15) \quad |\phi_\mu(\theta) \exp\{-\lambda(e^{i\theta} - 1)\} - \tilde{A}^{(r)}(\theta)| \leq k_4 n |\theta|^{r+1+\delta} (1 + (n|\theta|^2)^{r-1}),$$

where  $k_4 = k_4(K^{(n)}, K_1, \dots, K_{r+1})$ . This gives

$$\begin{aligned} \gamma_1 &= nk_4, \quad t_1 = r + 1 + \delta, \quad \gamma_2 = n^r k_4, \quad t_2 = 3r - 1 + \delta \\ \gamma &=, \quad \rho = 2\lambda/\pi^2, \quad \varepsilon = 0, \quad \text{and } \theta_0 = n^{-1/3} \end{aligned}$$

in Proposition 2.2, together with  $\eta = Ce^{-n^{1/3}\rho'}$  from the earlier bounds. Applying Corollary 2.3, and using the tail properties of the Poisson–Charlier measures (3.11), the theorem follows.  $\square$

A total variation bound of precisely the same order can also be deduced, by combining the arguments used for Propositions 2.2 and 2.4. Note that  $\phi_\mu$  is twice differentiable, because the  $X_j$  all have finite second moments, and that, as in Section 3, we need to take  $\chi(\theta) := \exp\{e^{i\theta} - 1 - i\theta\}$ .

**6.3. Analytic combinatorial schemes.** An extremely interesting range of applications is to be found in the paper of Hwang (1999). His conditions are motivated by examples from combinatorics, in which generating functions are natural tools. He works in an asymptotic setting, assuming that  $X_n$  is a random variable whose probability generating function  $R_n$  is of the form

$$R_n(z) = z^h (g(z) + \varepsilon_n(z)) e^{\lambda(z-1)},$$

where  $h$  is a non-negative integer, and both  $g$  and  $\varepsilon_n$  are analytic in a closed disc of radius  $\eta > 1$ . As  $n \rightarrow \infty$ , he assumes that  $\lambda \rightarrow \infty$  and that  $\sup_{z: |z| \leq \eta} |\varepsilon_n(z)| \leq K\lambda^{-1}$ , uniformly in  $n$ . He then proves a number of results describing the accuracy of the approximation of  $P_{X_n-h}$  by  $\text{Po}(\lambda + g'(1))$ .

Under his conditions, it is immediate that we can write

$$(6.16) \quad g(z) = \sum_{j \geq 0} g_j (z-1)^j \quad \text{and} \quad \varepsilon_n(z) = \sum_{j \geq 0} \varepsilon_{nj} (z-1)^j$$

for  $|z| < \eta - 1$ , with

$$(6.17) \quad |g_j| \leq k_g (\eta - 1)^{-j} \quad \text{and} \quad |\varepsilon_{nj}| \leq \lambda^{-1} k_\varepsilon (\eta - 1)^{-j}$$

for all  $j \geq 0$ . Hence  $X := X_n - h$  has characteristic function of the form  $\psi p_\lambda$ , where

$$\psi^{(n)}(\theta) = g(e^{i\theta}) + \varepsilon_n(e^{i\theta}),$$

and hence, for any  $r \in \mathbb{N}_0$ ,

$$(6.18) \quad |\psi^{(n)}(\theta) - \tilde{\psi}_r^{(n)}(\theta)| \leq K_{r1} |\theta|^{r+1}, \quad |\theta| \leq (\eta - 1)/2,$$

with  $\tilde{\psi}$  defined as in (3.4), taking  $\tilde{a}_j^{(n)} = g_j + \varepsilon_{nj}$ ; note that the constant  $K_{r1}$  can indeed be taken to be uniform for all  $n$ . Since also  $g$  and  $\varepsilon_n$  are both uniformly bounded on the unit circle, and since  $\tilde{\psi}_n$  is bounded (uniformly in  $n$ ) for  $|\theta| \leq \pi$ , it is clear that (6.18) can be extended to all  $|\theta| \leq \pi$ , albeit with a different uniform constant  $K'_{r1}$ , so that (3.3) holds with  $\delta = 1$  for any  $r \in \mathbb{N}_0$ . Thus Theorems 3.1 and 3.2 can be applied with any choice of  $r$ , giving progressively more accurate approximations to  $P_{X_n-h}$ , as far as the  $\lambda$ -order is concerned, in terms of progressively more complicated perturbations of the Poisson distribution. These theorems are thus applicable to all the examples that Hwang considers, including the numbers of components (counted in various ways) in a wide class of logarithmic assemblies, multisets and selections.

For instance, Corollary 4.2 gives an approximation to  $P_{X_n-h}$  by the mixture  $Q_{\lambda' mp}$  with

$$m := \lfloor m_n - v_n \rfloor; \quad p^2 := \langle m_n - v_n \rangle; \quad \lambda' := \lambda + v_n - p(1 - p),$$

where  $m_n := g'_n(1)$ ,  $v_n := g''_n(1) + g'_n(1) - \{g'_n(1)\}^2$  and  $g_n := g + \varepsilon_n$ . Hwang's approximation by  $\text{Po}(\lambda + g'(1))$  has asymptotically the same mean as ours (and as that of  $X_n - h$ ), but a variance asymptotically differing by  $\kappa := g''(1) - \{g'(1)\}^2$  (together with an element arising from  $p(1 - p)$  which is not in general asymptotically negligible). As a consequence, Hwang's approximation has an error of larger asymptotic order, in which the quantity  $\kappa$  appears; for instance, for Kolmogorov distance, his Theorem 1 gives an error of order  $O(\lambda^{-1})$ , whereas that from Corollary 4.2 is of order  $O(\lambda^{-3/2})$ .

Although our Poisson expansion theorems are automatically applicable under Hwang's conditions, they also apply to examples that do not satisfy his conditions: that of Section 6.1 is one such. Conversely, Hwang's Theorem 2, which establishes Poisson approximation in the lower tail with good *relative* accuracy, cannot be proved using only our conditions; the conclusion would not be true, for instance, for the random variable  $X - Y_s$  of Section 6.1.

Note also that Hwang examines problems from combinatorial settings in which approximation is not by Poisson distributions: he has examples concerning the Bessel family,

$$B(\lambda)\{j\} := L(\lambda)^{-1} \frac{\lambda^j}{j!(j-1)!}, \quad j \in \mathbb{N},$$

for the appropriate choice of  $L(\lambda)$ . Here, we could apply Corollary 5.2 to obtain slightly sharper approximations than his within the translated Bessel family, or Theorem 5.1 to obtain asymptotically more accurate expansions.

**6.4. Prime divisors.** The numbers of prime divisors of a positive integer  $n$ , counted either with  $(\Omega(n))$  or without  $(\omega(n))$  multiplicity, can also be treated by these methods, since excellent information is available about their generating functions. For our purposes, we use only the shortest expansion, taken from Tenenbaum (1995, Theorems II.6.1 and 6.2). One finds

that for  $N_n$  uniformly distributed on  $\{1, 2, \dots, n\}$  we have

$$\begin{aligned}\mathbb{E}\{e^{i\theta\omega(N_n)}\} &= p_{\log \log n}(\theta) \left\{ \Phi_1(e^{i\theta} - 1) + \eta_1(\theta) \right\}; \\ \mathbb{E}\{e^{i\theta\Omega(N_n)}\} &= p_{\log \log n}(\theta) \left\{ \Phi_2(e^{i\theta} - 1) + \eta_2(\theta) \right\},\end{aligned}$$

where  $|\eta_s(\theta)| \leq C_s / \log n$ ,  $s = 1, 2$ , for some constants  $C_1$  and  $C_2$ , and

$$\begin{aligned}\Phi_1(w) &:= \frac{1}{\Gamma(1+w)} \prod_q \left(1 + \frac{w}{q}\right) \left(1 - \frac{1}{q}\right)^w; \\ \Phi_2(w) &:= \frac{1}{\Gamma(1+w)} \prod_q \left(1 - \frac{w}{q-1}\right)^{-1} \left(1 - \frac{1}{q}\right)^w,\end{aligned}$$

$q$  running here over prime numbers. These expansions were established and used by Rényi and Turán (1958) in their proof of the Erdős–Kac Theorem, but they are also sketched by Selberg (1954). We refer to Kowalski and Nikeghbali (2009) for the structural interpretation of the two factors in these functions (with  $1/\Gamma(1+w)$  being related to the number of cycles of large random permutations).

Let  $\tilde{a}_{ls}$ ,  $s = 1, 2$ , denote the Taylor coefficients of the functions  $\Phi_s(w)$  as power series in  $w$  (around  $w = 0$ , which corresponds to  $\theta = 0$ ). By analyticity, it follows that for any  $r$ , we have

$$\left| \Phi_s(w) - 1 - \sum_{l=1}^r \tilde{a}_{ls} w^l \right| \leq C_{rs} |w|^{r+1},$$

for suitable constants  $C_{rs}$  and for  $|w| \leq 2$ . Defining the measures  $\nu_r^{(s)}$  by

$$\nu_r^{(s)}\{j\} := \text{Po}(\log \log n)\{j\} \left(1 + \sum_{l=1}^r (-1)^l \tilde{a}_{ls} C_l(j; \log \log n)\right),$$

this leads to the following conclusion, which is deduced immediately from Theorem 3.1, and refines the Erdős–Kac theorem.

**Theorem 6.2.** *For the measures  $\nu_r^{(s)}$  defined above, we have*

$$\begin{aligned}d_{\text{loc}}(P_{\omega(N_n)}, \nu_r^{(1)}) &\leq \alpha'_{1,r+1} C_{r1} (\log \log n)^{-1-r/2} + \tilde{a}_1 C_1 / \log n; \\ d_K(P_{\omega(N_n)}, \nu_r^{(1)}) &\leq \alpha'_{2,r+1} C_{r1} (\log \log n)^{-(r+1)/2} + \tilde{C}_1 \log \log n / \log n; \\ d_{\text{loc}}(P_{\Omega(N_n)}, \nu_r^{(2)}) &\leq \alpha'_{1,r+1} C_{r2} (\log \log n)^{-1-r/2} + \tilde{a}_1 C_2 / \log n; \\ d_K(P_{\Omega(N_n)}, \nu_r^{(2)}) &\leq \alpha'_{2,r+1} C_{r2} (\log \log n)^{-(r+1)/2} + \tilde{C}_2 \log \log n / \log n,\end{aligned}$$

for suitable constants  $\tilde{C}_1$  and  $\tilde{C}_2$ .

**Remark.** Note that it follows from Theorem 3.2 that the total variation distance is in each case also of order  $O\{(\log \log n)^{-(r+1)/2}\}$ . This can be deduced by applying the theorem to the expansion with one more term, and then observing that the extra term has total variation norm of order  $O\{(\log \log n)^{-(r+1)/2}\}$ , in view of the observation following (3.9). Alternatively, one could use Proposition 2.4. As far as we know, total variation approximation was first considered in this context by Harper (2009), who proved a bound with error of size  $1/(\log \log n)$  (for a truncated version of

$\omega(n)$ , counting only prime divisors of size up to  $n^{1/(3(\log \log n)^2)}$ , and deduced explicit bounds in Kolmogorov distance.

To indicate what this means in concrete terms for number theory readers, consider the case of  $\omega(n)$  for  $r = 1$ . Taylor expansion gives

$$\Phi_1(w) = 1 + B_1 w + O(w^2)$$

as  $w \rightarrow 0$ , where  $B_1 \approx 0.26149721$  is the Mertens constant, i.e., the real number such that

$$\sum_{\substack{q \leq x \\ q \text{ prime}}} \frac{1}{q} = \log \log x + B_1 + o(1),$$

as  $x \rightarrow +\infty$ .

In view of the remark above, an application of Theorem 6.2 gives

$$\begin{aligned} \left| \frac{1}{n} |\{k \leq n \mid \omega(n) \in A\}| - \nu_1^{(1)}\{A\} \right| &\leq \frac{1}{2} \|P_{\omega(N_n)} - \nu_1^{(1)}\| \\ &= O\left(\frac{1}{\log \log n}\right), \end{aligned}$$

for any set  $A$  of positive integers, where

$$\nu_1^{(1)}\{j\} = \text{Po}(\log \log n)\{j\} \left(1 + B_1 \left\{1 - \frac{j}{\log \log n}\right\}\right).$$

Higher expansions could be computed in much the same way.

Alternatively, a more accurate approximation is available from Corollary 4.2, while staying within the realm of (translated) Poisson distributions.

For this, we compute the expansion of  $\Phi_1$  to order 2, obtaining (after some calculations) that

$$\Phi_1(w) = 1 + B_1 w + \tilde{a}_2 w^2 + O(w^3), \quad \text{as } w \rightarrow 0,$$

where

$$\tilde{a}_2 := \frac{B_1^2}{2} - \frac{\pi^2}{12} - \frac{1}{2} \sum_{q \text{ prime}} \frac{1}{q^2}$$

(use  $1/\Gamma(1+w) = 1 + \gamma w + (\gamma^2 - \pi^2/12)w^2 + O(w^3)$ , as well as the Mertens identity

$$\gamma + \sum_{q \text{ prime}} \left( \frac{1}{q} + \log\left(1 - \frac{1}{q}\right) \right) = B_1,$$

and expand every term in the Euler product). This corresponds to (3.3), since  $w = e^{i\theta} - 1$ , and therefore we have (3.1) with

$$a_1 = B_1, \quad a_2 = \tilde{a}_2 + \frac{1}{2}B_1 = \frac{B_1 + B_1^2}{2} - \frac{\pi^2}{12} - \frac{1}{2} \sum_{q \text{ prime}} \frac{1}{q^2}.$$

We can then apply Corollary 4.2 to get the translated Poisson approximation  $Q_{\lambda mp}$ , with parameters calculated using (4.4). With

$$x := B_1 - (2a_2 - B_1^2) = \frac{\pi^2}{6} + \sum_{q \text{ prime}} \frac{1}{q^2} \approx 2.0971815,$$

this gives

$$\begin{aligned} p &= \sqrt{\langle x \rangle} \approx 0.31173945; & m &= 2; \\ \lambda' &= \log \log n + B_1 - x - p(1-p) \approx \log \log n - 2.0502422 \end{aligned}$$

Thus for any positive integer  $n$  and any set  $A$  of positive integers, we have

$$\begin{aligned} \left| \frac{1}{n} |\{k \leq n \mid \omega(n) \in A\}| - \{p\text{Po}(\lambda')\{A-3\} + (1-p)\text{Po}(\lambda')\{A-2\}\} \right| \\ = O\left(\frac{1}{(\log \log n)^{3/2}}\right), \end{aligned}$$

where, again, we can use the total variation norm in view of the previous remark. Similar results hold for  $\Omega(n)$ , where one obtains the following approximate values

$$\begin{aligned} p &\approx 0.5195; & m &= 0; \\ \lambda' &\approx \log \log n + 0.5152. \end{aligned}$$

#### REFERENCES

- [1] A. D. BARBOUR & V. ČEKANAČIUS (2002) Total variation asymptotics for sums of independent integer random variables. *Ann. Probab.* **30**, 509–545.
- [2] T. S. CHIHARA (1978) *An introduction to orthogonal polynomials*. Gordon and Breach, New York.
- [3] F. CHUNG & L. LU (2006) Concentration inequalities and martingale inequalities: a survey. *Internet Math.* **3**, 79–127.
- [4] A. J. HARPER (2009) Two new proofs of the Erdős–Kac Theorem, with bound on the rate of convergence, by Stein’s method for distributional approximations. *Math. Proc. Cam. Phil. Soc.* **147**, 95–114.
- [5] H.-K. HWANG (1999) Asymptotics of Poisson approximation to random discrete distributions: an analytic approach. *Adv. Appl. Prob.* **31**, 448–491.
- [6] J. JACOD, E. KOWALSKI & A. NIKEGHBALI (2008) Mod–Gaussian convergence: new limit theorems in probability and number theory, to appear in *Forum Math.*; see also [arXiv:0807.4739](#).
- [7] E. KOWALSKI & A. NIKEGHBALI (2009) Mod–Poisson convergence in probability and number theory. [arXiv:0905.0318](#).
- [8] V. V. PETROV (1975) *Limit theorems of probability theory*. Oxford University Press, Oxford.
- [9] A. RÉNYI & P. TURÁN (1958) On a theorem of Erdős–Kac. *Acta Arith.* **4**, 71–84.
- [10] A. SELBERG (1954) Note on the paper by L.G. Sathe. *J. Indian Math. Soc.* **18** 83–87.
- [11] G. TENENBAUM (1995) *Introduction à la théorie analytique et probabiliste des nombres*. Société Mathématique de France.

INSTITUT FÜR MATHEMATIK, UNIVERSITÄT ZÜRICH, WINTERTHURERTRASSE 190, CH-8057 ZÜRICH, SWITZERLAND

*E-mail address:* [a.d.barbour@math.uzh.ch](mailto:a.d.barbour@math.uzh.ch)

ETH ZÜRICH, D-MATH, RÄMISTRASSE 101, 8092 ZÜRICH, SWITZERLAND

*E-mail address:* [kowalski@math.ethz.ch](mailto:kowalski@math.ethz.ch)

INSTITUT FÜR MATHEMATIK, UNIVERSITÄT ZÜRICH, WINTERTHURERTRASSE 190, CH-8057 ZÜRICH, SWITZERLAND

*E-mail address:* [ashkan.nikeghbali@math.uzh.ch](mailto:ashkan.nikeghbali@math.uzh.ch)